

빅데이터 활용 Big Data 2

소프트웨어와 미래사회

2019



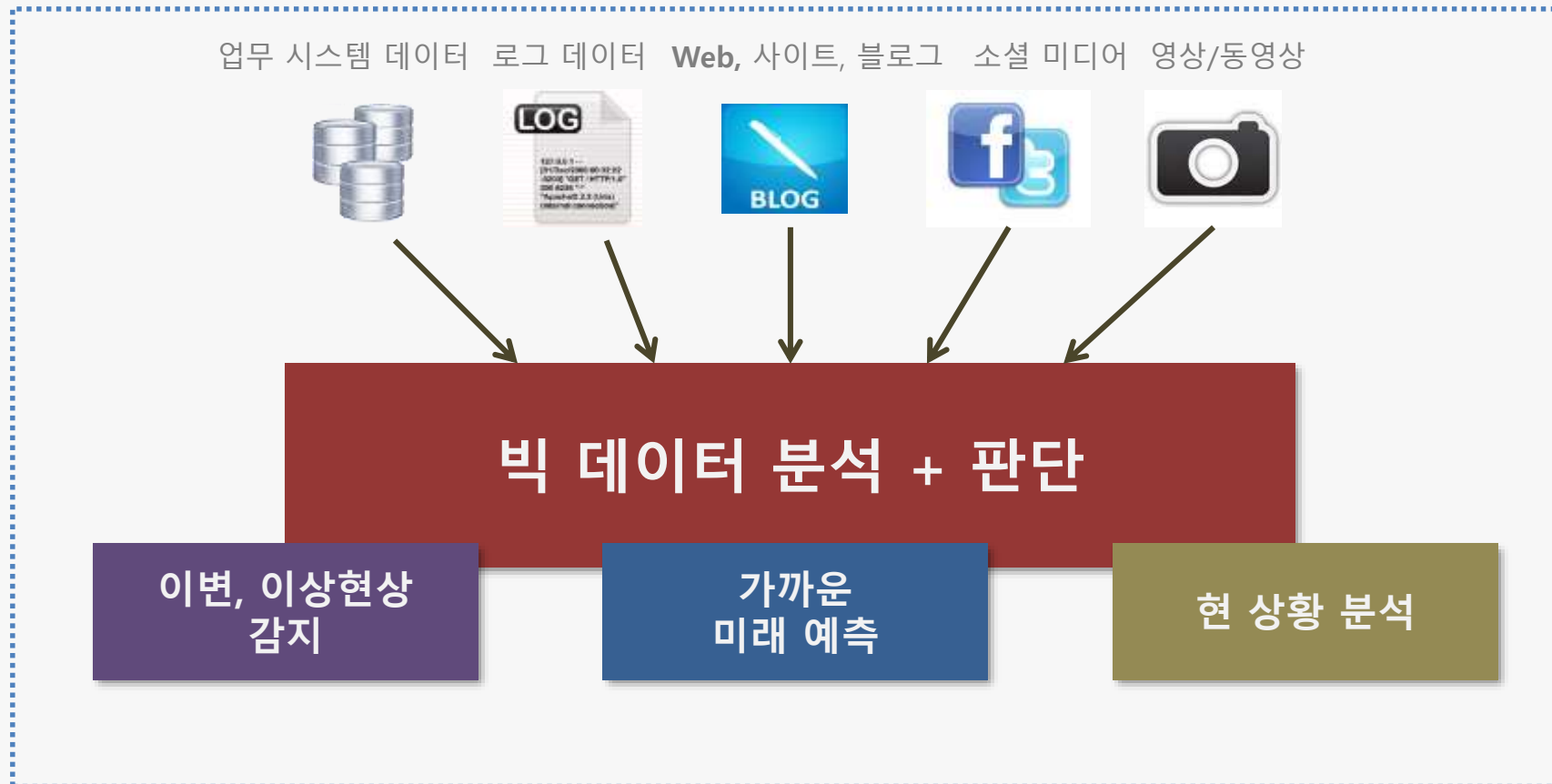
Big Data

미숙아 질병 감염 발견 빅데이터 - 세계미래포럼 (2014)

[빅데이터 활용분야]

○ 빅데이터로 할 수 있는 것들

- 다양하고 많은 분야에서 빅데이터를 활용



[빅데이터 활용사례 : 판매]

<파리바게트, 날씨지수에 따른 제품 판매 현황>



- 기상관측 자료와 10억 건 이상의 점포별 상품 판매 데이터를 분석
- 실시간으로 전국 파리바게뜨 점포의 단말기에 제공
 - 생크림 케익 항목 옆에 '토요일 50.15%, 일요일 27.15%' 같은 숫자 표기
 - 매장 주변 날씨예보, 요일 등의 요소를 종합해서 최근 2주 평균보다 매출이 증가함을 의미
- 날씨 지수를 도입한지 한달 만에 조리 빵 매출 30% 증가

[빅데이터 활용사례 : 범죄예방]

미국 샌프란시스코 : 범죄예방시스템으로 안전 지역사회 구축



- 과거범죄 기록에서 범죄 패턴을 분석하여
후속범죄의 발생지역 및 시각 예측
- 8년간 범죄 데이터 분석을 통한
범죄 발생 예측시스템 운영
(약 71% 정확도)

미국 FBI : 유전자 색인 시스템 활용한 단시간 범인 검거



- 1시간 내에 범인 DNA 분석을 위한 주정부
데이터 연계 및 빅데이터 실시간 분석
솔루션 확보를 통해 범죄사건 해결
- 내장된 DNA 분석정보를 활용하여
45,500건의 범인 DNA적중

[빅데이터 활용사례 : 장바구니 분석]

기저귀 옆에 맥주를 진열?



월마트

- 마루이이마이 백화점

스포츠 용품 매장에서 상품을 구입한 고객의 절반이 한 달 이내에 여성용 양말을 구입했다는 사실을 발견

➔ 고객들에게 DM을 발행 ➔ DM을 이용한 상품 구입율이 1% 미만에서 10% 넘는 수준으로 상승

[빅데이터 활용사례 : 스포츠]

- 독일 축구대표팀 – 2014 브라질 월드컵 준결승전
 - 많은 전문가들이 홈어드벤티지가 있는 브라질 대표팀 승리 예상
 - 독일 대표팀 7:1로 승리 후 아르헨티나 꺾고 우승 차지
- 독일축구협회 IT 기업 SAP사와 함께 독일 축구 대표팀 선수들 분석
 - 매치 인사이트" 라는 프로그램을 이용 선수들의 무릎, 어깨에 4개의 센서를 부착하고 훈련 혹은 경기 중 선수들의 체력, 맥박, 근육의 움직임 등 선수들의 경기 데이터를 수집
 - 축구공에도 센서를 부착, 공을 다루는 선수들의 속도, 습관등의 데이터 수집
 - 데이터를 바탕으로 1:1 훈련 및 악습관 교정
 - 평균 공 소유시간을 70% 줄여 빠른 공격 전개



[빅데이터 활용사례 : 고객분석]



미국 대형할인점 '타깃'



- 마트가 먼저 안 '엄마의 탄생'

- 데이터 분석 전문가 앤드루 폴 영입 → 임신 예측 모델 개발
- 임신 4~6개월 무렵부터 관련 상품 구매
- 쇼핑 습관이 그 어떤 시기보다 변하기 쉽다
- 시기에 맞춰 쿠폰과 인센티브 제공
- 베이비샤워 프로그램 이용
- 임신지수모델, 10대 딸 임신 알려
- 더 나아가 타깃은 쿠폰을 받은 임산부들의 반응에도 신경을 더 쓰기 시작

Big Data

[빅데이터 활용사례 : 대중교통]

- 서울시 올빼미 버스
 - 서울시와 KT는 MOU(Memorandum of Understanding)를 체결,
고효율 야간 버스 노선을 만들기위해 빅데이터를 사용
 - 서울시가 기존에 보유하고있던 교통 빅데이터와
KT의 시간/요일별 유동인구 빅데이터를 결합 각각의 시간별 인구를 파악
 - 서울시 전역을 1KM단위로 1252개 구역으로 나눈뒤,
지도위에 시간별 유동인구를 채워, 분석결과를 시각화
 - 모든 차량에 과속방지장치와 격벽을 설치
 - 운전자가 낮시간때 타업무를 하지못하도록
야간 버스 운전자들의 처우를 개선. 졸음방지



[빅데이터 활용사례 : 대중교통]

- 카카오택시

라스트 원 마일(last one mile) :

데이터 분석 결과 출퇴근 시간대 2~3km의 짧은 거리에서 택시 이용빈도가 높은 구간이 있다는 것을 발견
상암동 디지털미디어시티역 일대, 사당역 인근의 A사 오피스 빌딩

특징

- 지하철역에서 2~3km 내에 대형 오피스 빌딩이나 대규모 거주단지 존재
- 버스가 운행하는 구간이라도 도보로 10분 이상 이동 필요
- 택시를 이용하면 5분 이내



[카카오가 분석한 '3km 내 단거리 택시' 호출과 이동 경로]

[빅데이터 활용사례 : 패션]

- 빠르게 제품을 바꾸어 가는 **패스트 패션**
 - 1년을 15개 시즌으로 나누어 제품을 디자인
완성된 제품은 단 4시간 안에 재단, 포장, 출하의 과정을 거쳐
전 세계 2,200개 매장으로 직송
 - **고객의 불편이나 고객의 요구등을 근무 직원이 데이터 입력 → 비정형 데이터 생성**
 - **유니클로에도 이런 데이터를 뽑아내는 팀이 존재**

“우리에겐 고객과 시장에서 얻은 데이터가 있고
이를 기반으로 고객이 흥미를 느낄 수 있는
패션 아이템을 제공할 뿐이다”

- 자라의 모회사 인디텍스 회장 파블로 아일라(Pablo Isla)



[빅데이터 활용사례 : 부동산 시장]

- **프롭테크(PropTech)**

- 부동산(Property)과 기술(Technology)의 합성어
- 부동산 데이터를 이용한 스타트업 기업과 디지털화된 부동산 서비스 제공 기업
- 최근 임대·운영 관리, 감정평가, 중개자문, 종합자산관리 등
수요자 편의를 우선하는 서비스 산업으로 변화

- 부동산O2O 기업 채널브리즈(직방)와 스테이션3(다방)의 성장
- 아파트 매매로 영역을 넓힌 직방은 'VR(가상현실) 홈투어'에 이어
부동산 실거래가와 매물 데이터를 비교 분석하고
주변정보를 알려주는 '빅데이터랩'을 서비스

- 낮은 부동산 산업의 투명성
- 업권 간 겹침금지
- 정부출자 벤처펀드의 부동산 투자제한 등의 국내 한계



네이버 데이터랩 <http://datalab.naver.com>

- 실시간 검색어, 분야별 검색어 확인 가능
- 성별, 연령별 검색어 트렌드 분석 가능

네이버 광고 <http://searchad.naver.com>



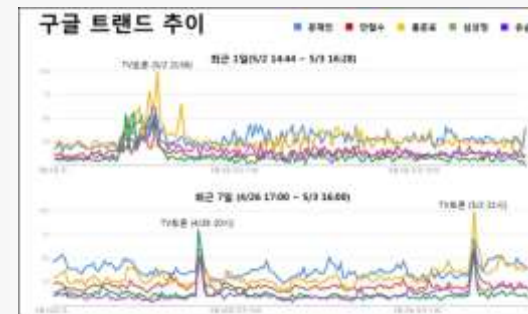
[빅데이터 활용사례 : 일상 속 빅데이터 활용]

카카오 키워드광고 <https://clix.biz.daum.net>

- 네이버 광고 - 광고시스템 - 도구 - 키워드 도구
- 카카오키워드 광고 - 광고관리 - 키워드 맞춤제안

구글 트렌드 <https://trends.google.com/trends>

- 전세계 검색
- 카카오키워드 광고 - 광고관리 - 키워드 맞춤제안



[빅데이터 활용 : 기타]

구글 번역기

최대 103개의 언어 번역, 책과 검색어 등의 데이터를 바탕으로 단시간에 고품질 번역 가능

미국 워싱턴 D.C.적십자사 본부

통신량이 증가하면 지도에 풍선모양이 커지는 대형 스크린 설치.
갑자기 부풀어 오르는 풍선은 사건이 있을 가능성이 높은 신호

지오프로스

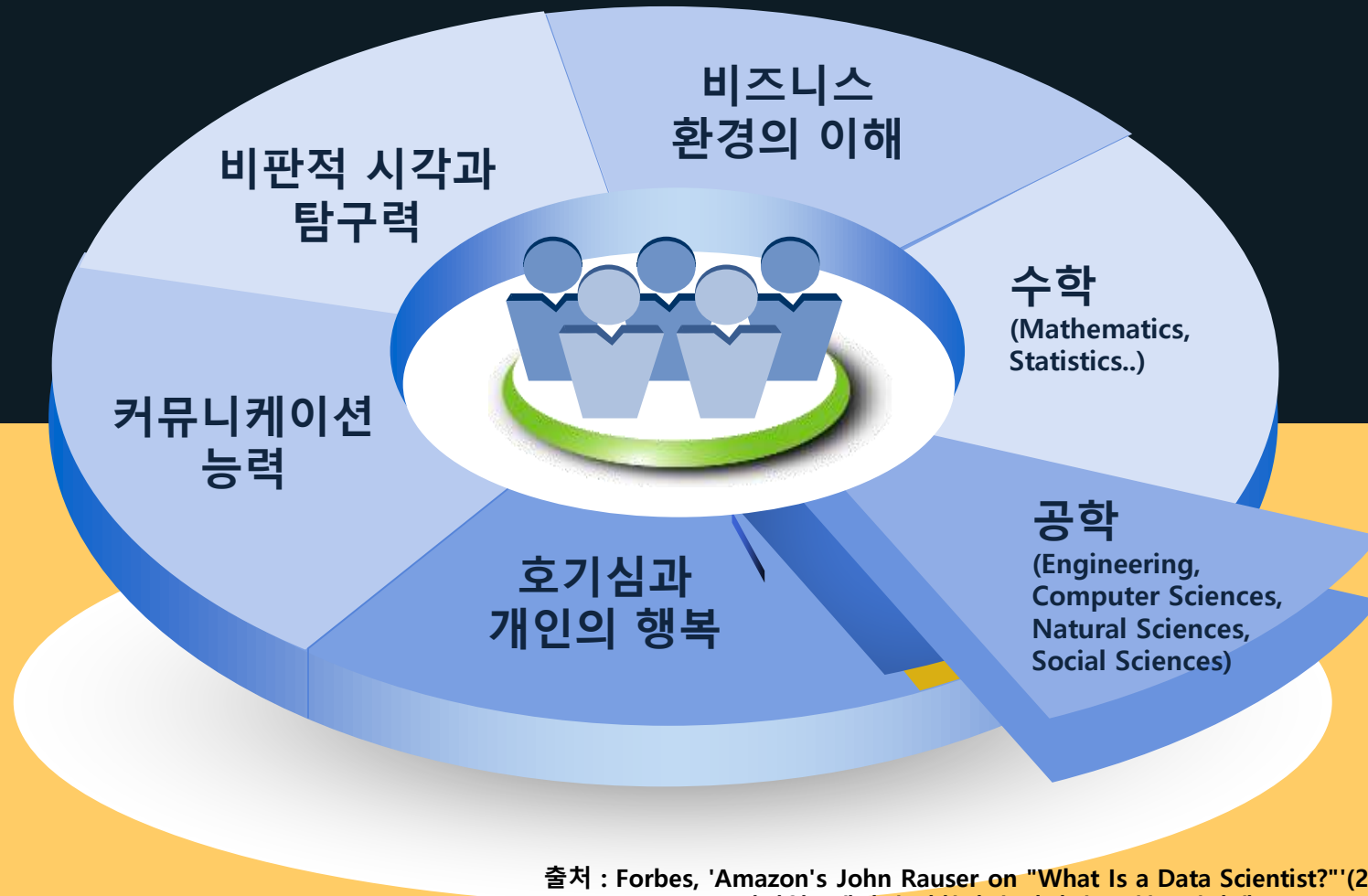
우리나라 경찰청에서 사용하는 지리적 프로파일링 프로그램
그동안 발생한 범죄장소와 행동반경 등을 분석해 범인의 범행 예측

패스트 데이터 : 데이터 중에서도 빠르게 확인 할 수 있는 데이터
(안심귀가 서비스, 운송 경로 및 상태 파악, 병원 진료 등)

스마트 데이터 : 빅데이터에서 찾아낸 꼭 필요한 정보

[빅데이터 활용을 위한 핵심 요소]

빅데이터 전문 인력의 자질



출처 : Forbes, 'Amazon's John Rauser on "What Is a Data Scientist?"'(2011.10.7), 이지영, '데이터 과학자가 되려면...', 블로터닷넷(2012. 3. 18) 재인용

[데이터 관련 용어]

- 데이터 스크래핑 Data scrapping

- 고객이 자신의 인증 정보를 한 번만 제공해도 컴퓨터가 금융기관, 대법원 등 시스템에 접속해 여러 곳에 흩어져 있는 고객의 정보를 모아 가공하거나 제공하는 기술
- بانک샐러드, 브로콜리 등의 모바일 가계부 서비스 등장
- 정부에서는 해킹으로 인한 피해 발생 이유로 해당 기술을 제한하는 방향 검토

- 마이데이터 (My Data) 산업

- 개인이 직접 금융기관과 통신사, 병원 등에 있는 자신의 개인정보를 제3의 업체에 전달해 새로운 서비스를 받도록 하는 정부의 시범 사업
- '데이터 이동권' 개념



출처 : 금융위원회

[10대 데이터&분석 기술 트렌드]

글로벌 IT 자문기관 가트너 발표(2019)

1. 증강분석

- 기업에서 발생한 머신러닝 기술을 이용하여 미래를 예측, 대안을 제시하는 기술
- 기존의 비즈니스 인텔리전스(BI)의 한계를 극복

2. 증강데이터 관리

- 머신러닝과 AI엔진을 활용하여 기업 정보 관리 카테고리 생성
- 데이터 품질, 메타데이터관리, 데이터통합, DBMS 자동 구성 등을 동적 시스템으로 전환
- 관리 수작업이 약 45% 줄어들 것으로 전망

3. 설명 가능한 AI

- 고급 AI 의 대다수가 권고사항이나 결정에 도달한 이유를 설명하지 못함
- 정확성, 속성, 모델 통계 기능 측면에서 모델에 대한 설명을 자연어로 자동 생성



[10대 데이터&분석 기술 트렌드]

해외 시장조사 및 컨설팅 업체, 매년 10월을 전후해 기술 영역에서 열 가지 핵심 주제 발표

4. 지속적 지능화

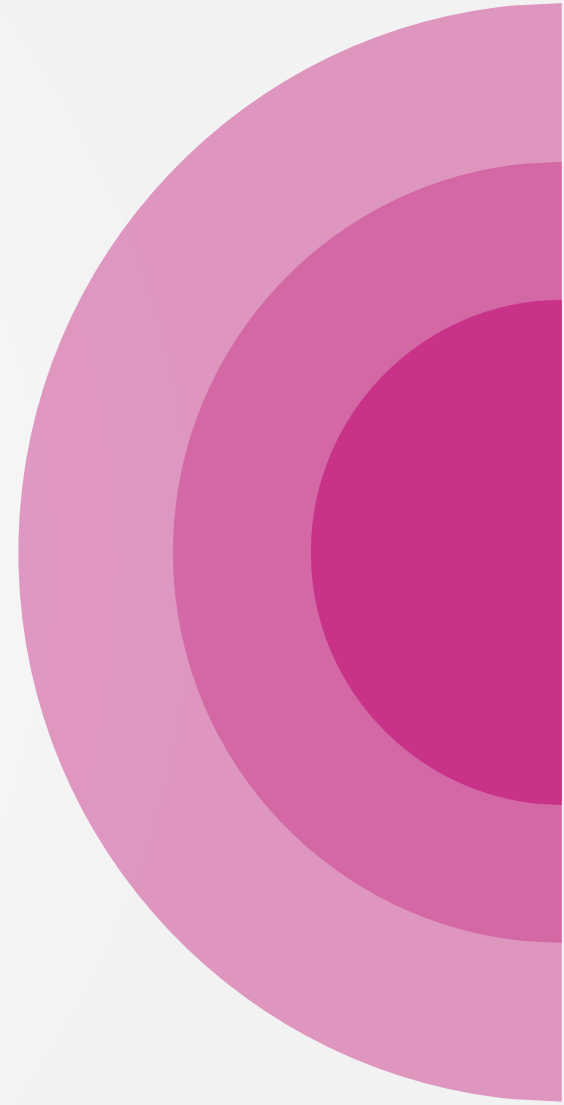
- 대용량 데이터를 예측에 활용해 실시간으로 대응하는 기술
- 현재와 과거의 데이터를 처리에 사건에 대응하는 자동화된 의사 결정 또는 지원

5. 그래프 분석

- 조직, 사람, 거래 등 주체간 관계를 탐색할 수 있는 분석 기법
- 복잡하고 적응력이 뛰어난 데이터 사이언스 구현 가능

6. 데이터 패브릭

- 분산된 데이터 환경에서 마찰 없는 Access와 데이터 공유 가능하게 함
- 일관된 단일 데이터 관리 프레임 워크 구축



[10대 데이터&분석 기술 트렌드]

7. 자연어처리 NLP

- 대화형 분석을 통한 용이한 접근

8. 상용 AI

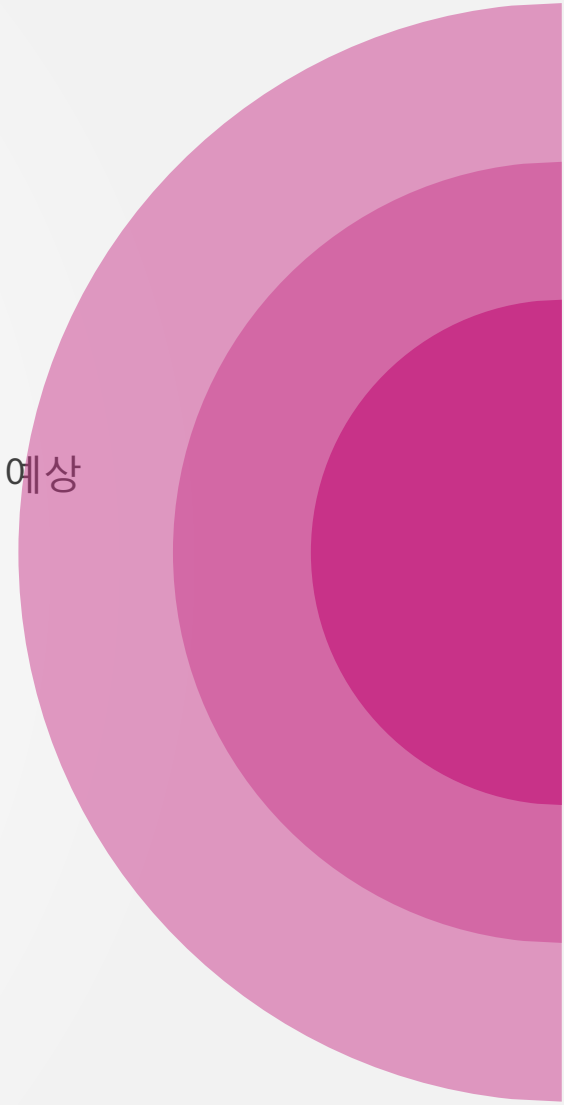
- 2022년에 이르면 최종 사용자 75%가 오픈 소스 플랫폼이 아닌 상용 솔루션으로 구축 예상

9. 블록체인

- 신뢰할 수 없는 참여자 네트워크 전반에 신뢰를 제공
- 기존 데이터 및 분석 인프라 통합

10. 퍼시스턴트 메모리 서버

- 퍼시스턴트 메모리는 고성능 워크로드에 비용 효율적인 대용량 메모리를 제공





THANK YOU
FOR LISTENING!